บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

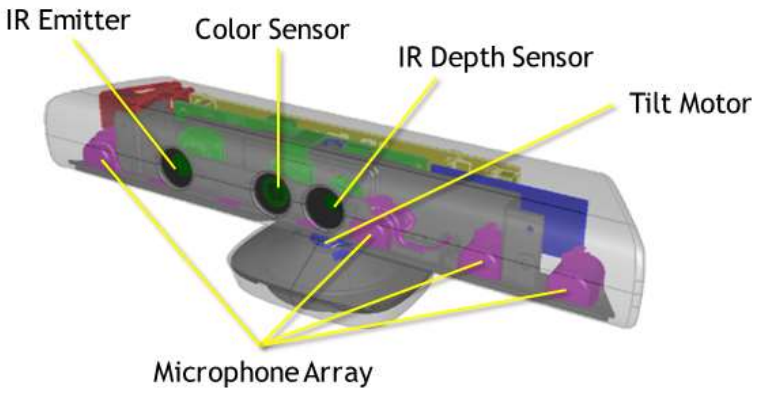
2.1 กล้องคินเน็คท์(Kinect)

คินเน็คท์คืออุปกรณ์เสริมสำหรับเครื่องเล่นเกมเครื่องเอ็กบ็อกซ์ 360 ซึ่งสามารถเชื่อมต่อกับคอมพิวเตอร์ผ่านทางพอร์ตยูเอสบี (USB Port) ทำให้สามารถรับข้อมูลจากตัวกล้องกับคอมพิวเตอร์ได้

รูปที่ 2.1 อุปกรณ์คินเน็คท์

2.1.1 องค์ประกอบของคินเน็คท์

1) Color Sensor 1 ตัว ทำหน้าที่เสมือนกล้องRGB เก็บข้อมูลความละเอียดสูงสุด 1280x960 ใช้สำหรับจับภาพสี
2) IR (Infrared) Emitter 1 ตัว ทำหน้าที่กระจายลำแสงอินฟราเรดแบบ Speckle ไปยังวัตถุ และมี IR Depth Sensor 1 ตัว ทำหน้าที่เป็นตัวรับอินฟราเรดที่สะท้อนกลับมายังเซ็นเซอร์ และจะแปลงค่าที่ได้เป็นค่าระยะความลึกระหว่างวัตถุและเซ็นเซอร์โดยจะอยู่ในหน่วยมิลลิเมตร
3) ไมโครโฟนมัลติอาร์เรย์ (Multi-array Microphone) ประกอบด้วยไมโครโฟน 4 ตัวสำหรับรับข้อมูลเสียงและสามารถค้นหาแหล่งกำเนิดเสียงจากคลื่นเสียงได้
4) มีมอเตอร์ ไว้สำหรับปรับมุมองศาของกล้องในแนวแกนตั้งได้



รูปที่ 2.2ส่วนประกอบต่างๆของกล้องคินเน็คท์

2.1.2 คุณสมบัติของคินเน็คท์

คินเน็คท์มีคุณสมบัติดังตารางที่ 2.1

ตารางที่ 2.1 คุณสมบัติของคินเน็คท์

คุณสมบัติ	**รายละเอียดคุณสมบัติ**
เซ็นเซอร์	ประมวลผลวิดีโอที่ได้ 30 เฟรมต่อวินาทีใช้งานได้ในระยะ 0.8 - 4 เมตร
มุมมอง	57.5องศาในแนวนอน 43 องศาในแนวตั้ง
อาร์จีบีวิดีโอสตรีม	ความละเอียดสูงสุดที่ระดับที่ (1280 x960พิกเซล) และที่ระดับวีจีเอ (640 × 480 พิกเซล)
มอเตอร์	หมุนเป็นมุมเอียงได้ 27 องศาในทิศขึ้นหรือลง
ไมโครโฟนมัลติอาร์เรย์	ประมวลผลแต่ละช่องเท่ากับ 16 บิตที่อัตราส่วนความถี่สุ่ม (Sampling Rate) เท่ากับ 16 กิโลเฮิร์ท
คินเน็คท์เรสซิเด้นท์	กำจัดเสียงก้องและกำจัดเสียงรบกวนออก
สั่งงานด้วยเสียง	คล้ายๆ Speech Recognition บน WindowsVista, Windows 7 โดยผู้ใช้สามารถใช้เสียงของผู้ใช้ในการสั่งงานให้ XBOX 360 สามารถทำสิ่งต่างๆได้
ใช้มือแทนเมาส์	ความสามารถในการตรวจจับการเครื่องไหวร่างกายของผู้เล่นทำให้ผู้ใช้สามารถใช้มือของคุณในการสั่งงานต่างๆบนหน้าจอได้

คุณสมบัติ	รายละเอียดคุณสมบัติ
จดจำลักษณะใบหน้า	Kinect สามารถจดจำใบหน้าของผู้เล่นได้โดยความสามารถนี้จะช่วยให้การเล่นเกมส์สนุกขึ้น
เชื่อมต่ออินเตอร์เน็ต	เพื่อความบันเทิงในการท่องเว็บต่างๆจากเครื่องเล่นเกมส์เครื่องเดียว

2.1.3 การทำงานของคินเน็คท์

กล้อง Kinect เป็นอุปกรณ์เสริมการใช้งานเครื่องเล่นเกม Xbox 360 มีความสามารถในการตรวจจับผู้เล่นและสามารถตรวจจับตำแหน่งต่างๆของร่างกายผู้เล่นได้โดยให้ข้อมูลตำแหน่งที่ตรวจจับได้เป็นแบบสามแกน (กว้าง x สูง x ลึก)ดังนั้นจึงสามารถนำมาประยุกต์ใช้กับการตรวจจับการเคลื่อนไหวของร่างกายแบบปราศจากการทำเครื่องหมายตามตำแหน่งต่างๆบนร่างกายได้ (Markerless Motion Capture) เมื่อใช้กล้อง Kinect ในการตรวจจับการเคลื่อนไหวของร่างกายข้อมูลตำแหน่งต่างๆของร่างกายที่ได้จะมีความถูกต้องเฉพาะส่วนที่กล้องตรวจจับได้หากมีบางส่วนของร่างกายถูกบดบังข้อมูลตำแหน่งของส่วนนั้นจะเป็นค่าประมาณมีความถูกต้องน้อยสามารถแก้ไขได้โดยการเพิ่มมุมกล้องในการตรวจจับซึ่งเป็นการเพิ่มจำนวนกล้องเพื่อทดแทนข้อมูลตำแหน่งที่บางมุมกล้องตรวจจับไม่ได้ด้วยอีกมุมกล้องหนึ่งทำให้ทุกส่วนของร่างกายถูกตรวจจับได้ตลอดเวลาแต่การทำเช่นนี้จะทำให้ได้ข้อมูลตำแหน่งต่างๆของร่างกายมากกว่าหนึ่งชุดในการตรวจจับคนหนึ่งคนจึงจำเป็นต้องนำข้อมูลตำแหน่งที่ได้มาประมวลผลรวมกันเพื่อให้ได้ข้อมูลตำแหน่งต่างๆของร่างกายที่มีความถูกต้องสมบูรณ์สามารถนำไปใช้ในการตรวจจับท่าทางและการเคลื่อนไหวของร่างกายต่อไปกล้องแต่ละกล้องที่ใช้ในการตรวจจับการเคลื่อนไหวของร่างกายในมุมมองที่แตกต่างกันจะทำให้ข้อมูลตำแหน่งต่างๆของร่างกายที่ได้มีตำแหน่งอ้างอิงต่างกันในการรวมข้อมูลตำแหน่งให้เป็นข้อมูลชุดเดียวกันหากรู้ทิศทางที่ตรวจจับและตำแหน่งของกล้องแต่ละตัวการรวมข้อมูลตำแหน่งจะสามารถทำได้โดยการหมุนข้อมูลตำแหน่งที่ได้ให้อยู่ในมุมกล้องทิศทางเดียวกัน (Rotation) แล้วทำการย้ายข้อมูลตำแหน่งให้อยู่ในตำแหน่งอ้างอิงเดียวกัน (Translation) จะทำให้ได้กลุ่มข้อมูลตำแหน่งที่ใกล้เคียงกันสามารถนำไปประมวลผลเพื่อเลือกใช้ข้อมูลตำแหน่งต่างๆของร่างกายต่อไปในการตรวจจับการเคลื่อนไหวของร่างกายหากใช้กล้องKinect หลายตัวในการตรวจจับการวัดหาตำแหน่งและทิศทางที่ตรวจจับของกล้องแต่ละตัวในพื้นที่จริงเป็นสิ่งยุ่งยากและไม่สะดวกงานวิจัยนี้จึงเสนอวิธีการวิเคราะห์หาตำแหน่งและทิศทางของกล้อง Kinect แบบอัตโนมัติโดยใช้ข้อมูลตำแหน่งของข้อเท้าที่กล้องแต่ละตัวตรวจจับได้ในการวิเคราะห์หาตำแหน่งและทิศทาง

Kinect มีหน้าที่หลักอยู่ 3 หน้าที่โดยสามารถเชื่อมระหว่าง Hardware กับหน้าที่การทำงานของ Kinect ดังนี้

1) จดจำผู้เล่นโดยใช้ข้อมูลจากกล้อง RGB (ข้อมูลแดงเขียวน้ำเงิน) และประมวลผลโดยวิธี FacialRecognition
2) จดจำการเคลื่อนไหวร่างกายของผู้เล่นเป็นแบบ 3 มิติแบ่งเป็น 3 อย่างย่อย
   2.1) จดจำวัตถุแบบ 3 มิติโดยใช้ตัวส่งแสง IR และกล้อง IRโดยตัวส่งแสง IR จะส่งแสงไปกระทบกับวัตถุและจะสะท้อนจากวัตถุนั้นๆกลับไปที่กล้อง IR ซึ่งKinect จะใช้ข้อมูลเวลาในการสะท้อนกลับและความยาวคลื่นแสงเพื่อประมวลผลระยะห่างระหว่างกล้องกับวัตถุ (ยกตัวอย่างเช่น ถ้าใช้เวลาในการสะท้อนกลับนานแสดงว่าวัตถุอยู่ไกล)ส่วนวัตถุที่มีสีต่างกันจะดูดกลืนและสะท้อนแสงกลับไปด้วยความยาวคลื่นที่ต่างกัน
   2.2) จดจำและแบ่งแยกประเภทของคนว่าเป็นเพศใด อายุประมาณเท่าไร ขนาดสัดส่วนตัวประมาณเท่าใด โดยใช้เทียบกับฐานข้อมูลที่มีอยู่
   2.3) จดจำการเคลื่อนไหวของคนโดยใช้วิธี Skeletal Movements(การเคลื่อนไหวตามลักษณะกระดูก)
3) จดจำเสียงผู้เล่นโดยใช้ไมโครโฟน 4 ตัวโดยไมโครโฟนนั้นจะเป็นแบบ Wide-Field, Conic Audio Capture รับเสียงในพื้นที่กว้างและให้ความสำคัญกับเสียงเป็นรูปกรวย (ตรวจจับเสียงไกลๆเสียงใกล้ๆ โดยยกเว้นเสียงที่มาจาก Xbox) หลังจากนั้นจะใช้วิธี Voice Recognition เพื่อจดจำและแปลความหมายของสิ่งที่ผู้เล่นแต่ละคนพูดโดย Kinectสามารถแยกแยะเสียงของแต่ละผู้เล่นได้

2.2 คินเน็คท์วินโดวส์เอสดีเค (Kinect for Windows SDK)

Kinect™ for WindowsSoftware Development Kit (SDK) จากไมโครซอฟท์เป็นไลบรารี่(Library) ที่ช่วยให้พัฒนาแอพพลิเคชันของคินเน็คท์บนระบบปฏิบัติการวินโดวส์ได้ง่ายขึ้น

2.2.1 ความต้องการของระบบ (System Requirememt)

จะต้องเรียกใช้โปรแกรมวินโดวส์เอสดีเคบนระบบปฏิบัติการหลักเท่านั้นไม่สามารถเรียกใช้บนโปรแกรมเวอร์ชวลแมชชีน (Virtual Machine) เนื่องจากไดร์ฟเวอร์ของคินเน็คท์ต้องติดตั้งบนคอมพิวเตอร์ที่ใช้งานเท่านั้น

2.2.1.1 ระบบปฏิบัติการและสถาปัตยกรรมที่รองรับ

1) Microsoft Windows 7 (x86 หรือ x64)
2) Windows Embedded Standard 7
3) กรณีระบบปฏิบัติการคือ Windows 7 N หรือ Windows 7 KN ต้องติดตั้ง Media Feature Pack สำหรับวินโดวส์รันไทม์

2.2.1.2 Hardware Requirements

1) ซีพียู (CPU) Dual-Core 2.66 GHz หรือมากกว่า
2) Windows 7 Compatible Graphics Card ที่รองรับ Microsoft DirectX 9.0c Capabilities
3) แรม (RAM) 2 GB
4) Kinect for Xbox 360 Sensor จะต้องมีสาย USB power ให้มาด้วย

2.2.1.3 Software Requirement

1) Microsoft Visual Studio. 2010 Express หรือ Visual Studio 2010 Edition อื่น
2) Microsoft .NET Framework 4.0 (ติดตั้งด้วย Visual Studio 2010)
3) DirectX Software Development Kit
4) DirectX End-User Runtime (June 2010)
5) Microsoft Speech Platform-Server Runtime เวอร์ชัน 11 (x86 Edition)
6) Microsoft Speech Platform-Software Development Kit เวอร์ชัน 11 (x86Edition)

2.2.2 การติดตั้งเอสดีเคและชุดเครื่องมือ (Toolkit)

2.2.2.1 การเตรียมความพร้อมการติดตั้งเอสดีเค

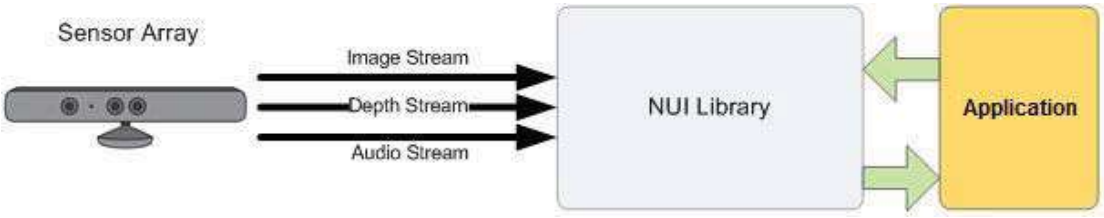
1) ตรวจสอบให้แน่ใจว่าอุปกรณ์คินเน็คท์ไม่ได้เชื่อมต่อกับเครื่องคอมพิวเตอร์
2) ตรวจสอบให้แน่ใจว่าได้ลงโปรแกรมตามที่ความต้องการระบบกำหนดไว้เรียบร้อยแล้ว
3) กรณีที่มีการติดตั้งเอสดีเครุ่นก่อนหน้าให้ถอดไดร์ฟเวอร์ที่เคยติดตั้งทุกตัวของอุปกรณ์คินเน็คท์
4) ปิด Visual Studio ก่อนการติดตั้งเอสดีเคและเริ่มระบบใหม่หลังจากการติดตั้งเสร็จ

2.2.2.2 ขั้นตอนการติดตั้งเอสดีเค

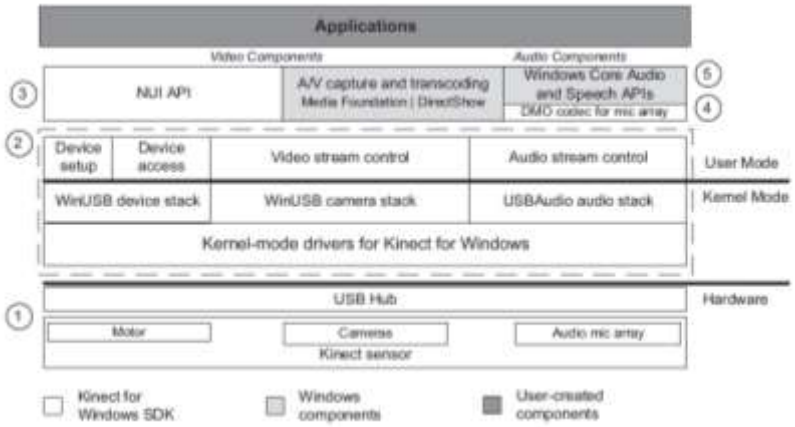
1) ในหน้าดาวน์โหลดจากเว็บไซต์คินเน็คท์ คลิกดาวน์โหลดKinectSDK-v1.5-Setup.exe ซึ่งสามารถติดตั้งได้ทั้งx86 และ x64
2) เมื่อเอสดีเคติดตั้งสมบูรณ์แล้วตรวจสอบให้มั่นใจว่าอุปกรณ์คินเน็คท์ได้ทำการเชื่อมต่อกับแหล่งจ่ายไฟภายนอกรวมถึงเชื่อมต่อกับเครื่องคอมพิวเตอร์ผ่านยูเอสบีไดร์ฟเวอร์จะมีการติดตั้งอย่างอัตโนมัติ
3) ติดตั้งชุดเครื่องมือ KinectDeveloperToolkit-v1.5.0-Setup.exe ซึ่งสามารถทำการติดตั้งได้ทั้ง x86 และ x64
4) หากต้องการสั่งการทำงานชุดอุปกรณ์หรือชุดคำสั่งตัวอย่าง ให้ติดตั้งชุดคำสั่งตัวอย่างหรือทรัพยากรแหล่งข้อมูลอื่นๆและทำการเริ่มต้น Developer Toolkit Browser ผ่านเมนูstart: Kinect for Windows SDK 1.5-> Developer Toolkit Browser v1.5.0 (Kinect for Windows)

2.2.3 สถาปัตยกรรมการออกแบบคินเน็คท์เพื่อใช้งานบนวินโดวส์

คินเน็คท์วินโดวส์เอสดีเคมีซอฟต์แวร์ไลบรารีที่ซับซ้อนและเครื่องมือที่ช่วยให้ผู้พัฒนานำข้อมูลจากคินเน็คท์มาใช้ได้ซึ่งเป็นข้อมูลที่ตอบสนองต่อเหตุการณ์จริงเซ็นเซอร์ของคินเน็คท์และไลบรารีของซอฟต์แวร์ที่เกี่ยวข้องในการโต้ตอบระหว่างแอพพลิเคชันกับตัวผู้ใช้แสดงในรูปที่ 2.4 และสถาปัตยกรรมการออกแบบของคินเน็คท์วินโดวส์เอสดีเคแสดงในรูปที่ 2.5



รูปที่ 2.3ฮาร์ดแวร์และซอฟต์แวร์ที่เกี่ยวข้องในการโต้ตอบระหว่างแอพพลิเคชัน



รูปที่ 2.4 สถาปัตยกรรมการออกแบบของคินเน็คท์วินโดวส์เอสดีเค

องค์ประกอบของสถาปัตยกรรมการออกแบบของคินเน็คท์วินโดวส์เอสดีเคได้แก่

1) Kinect Hardware คือส่วนประกอบฮาร์ดแวร์นับตั้งแต่คินเน็คท์เซ็นเซอร์และUSB Hubไปจนถึงเซ็นเซอร์ที่เชื่อมต่อกับคอมพิวเตอร์
2) Kinect Drivers คือวินโดวส์ไดร์ฟเวอร์สำหรับคินเน็คท์ที่ถูกติดตั้งไว้เป็นส่วนหนึ่งของกระบวนการติดตั้งเอสดีเค
3) Kinect Natural User Interface สำหรับการจับโครงร่างเสียงภาพสีและข้อมูลความลึกภาพ
4) DirectX Media Object (DMO) คือส่วนที่เพิ่มเติมจากไมโครโฟนอาร์เรย์ที่รองรับระบบปฏิบัติการวินโดวส์เพื่อใช้สำหรับส่งสัญญาณผ่านอากาศ (Beamforming) และสามารถแปลความหมายจากแหล่งที่มาได้
5) Windows 7 standard APIs คือเสียงคำพูดและสื่อ APIs บน Windows 7

2.2.4 Natural User Interface for Kinect for Windows

เป็นส่วนสำคัญของส่วนต่อประสานโปรแกรมประยุกต์ของคินเน็คท์สำหรับวินโดวส์ (Kinect for Windows API) รองรับการประมวลผลภาพขั้นพื้นฐานและการจัดการใช้งานอุปกรณ์ต่างๆมีดังต่อไปนี้

1) การเข้าถึงเซ็นเซอร์คินเน็คท์ที่เชื่อมต่อกับคอมพิวเตอร์ได้แก่ข้อมูลสตรีมเสียง ข้อมูลสตรีมภาพสีและสตรีมของข้อมูลความลึก

2) ซอฟต์แวร์ไปป์ไลน์ที่สามารถจดจำและตรวจจับโครงร่างมนุษย์โดยแปลงข้อมูลความลึกเป็นโครงร่างข้อต่อในร่างกายมนุษย์ทำให้สามารถตรวจจับสองโครงร่างมนุษย์ภายในเวลาเดียวกัน
3) เมื่อใช้ร่วมกับส่วนต่อประสานโปรแกรมประยุกต์ของไมโครซอฟท์สปีช (Microsoft Speech APIs) ทำให้สามารถจดจำคำพูดทำให้สามารถใช้คำสั่งเสียงได้
4) เมื่อใช้ร่วมกับเอสดีเคของการตรวจจับใบหน้า (Face Tracking SDK) ทำให้สามารถตรวจจับใบหน้ามนุษย์ได้

2.2.5 เอ็นยูไออิมเมจดาต้าสตรีม (NUI Image Data Streams)

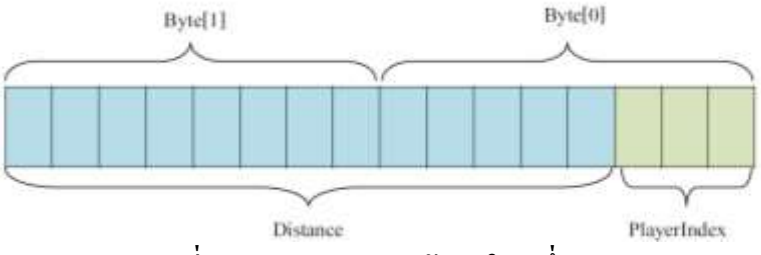
2.2.5.1 ข้อมูลความลึก (Depth Stream)

แต่ละเฟรมของข้อมูลความลึกถูกจัดอยู่ในรูปแบบพิกเซลประกอบด้วยระยะความลึกในหน่วยมิลลิเมตรที่ได้มาจากการทำงานของตัวส่งแสงและเซ็นเซอร์วัดระยะความลึกจากแสงอินฟราเรดระยะความลึกดังกล่าวคือระยะความลึกจากพื้นผิววัตถุที่อยู่ใกล้ที่สุดถึงระนาบกล้องซึ่งแอพพลิเคชันสามารถประมวลผลข้อมูลความลึกเพื่อตรวจจับการเคลื่อนที่ของบุคคลหรือระบุพื้นหลังภาพเพื่อแบ่งข้อมูลที่ส่งมามีความละเอียดได้ 3 ขนาดคือ

1) ขนาดความละเอียด 640x480 (มาตรฐาน)
2) ขนาดความละเอียด 320x240
3) ขนาดความละเอียด 80x60

สตรีมของข้อมูลความลึกประกอบด้วย 2 ประเภทรวม 16 บิตได้แก่

1) สตรีมของข้อมูลความลึก 13 บิต
2) สตรีมของข้อมูลการแบ่งส่วนผู้เล่นจำนวน 3 บิตเป็นบิตที่ต่อท้าย
3) สตรีมข้อมูลความลึกเป็นค่าจำนวนเต็มจากแผนผังการแบ่งแยกและไม่ถูกใช้เป็นแฟลกซ์ในฟิลด์บิต



รูปที่ 2.5ลักษณะการเก็บข้อมูลในหนึ่งพิกเซล

2.2.5.2 ข้อมูลการแบ่งแยกผู้เล่น (Player Segmentation Data)

คินเน็คท์ประมวลผลข้อมูลสีและความลึกเพื่อใช้ในการระบุโครงร่างมนุษย์ได้มากที่สุด ถึงหกโครงร่างในแผนผังการแบ่งส่วนแผนผังการแบ่งส่วนเป็นบิตแมปที่มีค่าพิกเซลสอดคล้องกับ Index ของบุคคลที่อยู่ในขอบเขตมุมมองที่ใกล้เคียงกับกล้องในตำแหน่งพิกเซลนั้นข้อมูลการแบ่งส่วนผู้เล่นจะใช้ได้ในสตรีมข้อมูลความลึกเมื่อส่วนการจับโครงร่างถูกเปิดใช้งาน

ข้อมูลการแบ่งส่วนผู้เล่นมักถูกเรียกโดยปกติว่าข้อมูลอินเด็กซ์ของผู้เล่นถึงแม้ว่าข้อมูลการแบ่งส่วนผู้เล่นเป็นสตรีมที่แบ่งแยกคือข้อมูลความลึก (13 บิต) และข้อมูลการแบ่งส่วน (3บิต) รวมเป็น 16 บิตของข้อมูลในแต่ละเฟรม

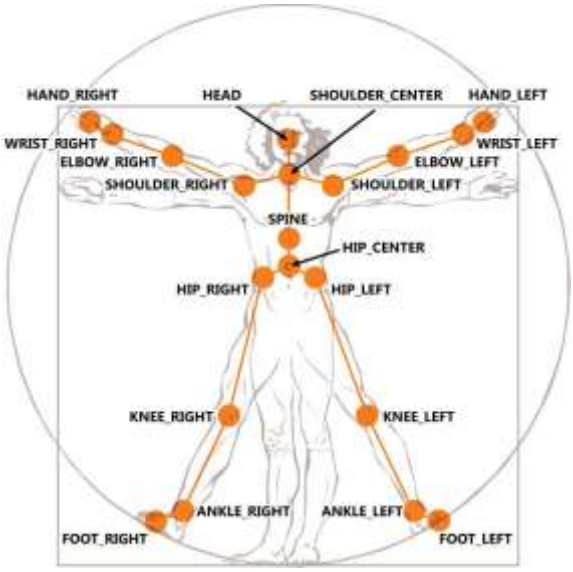
ค่า 0 แสดงให้เห็นว่าไม่มีผู้ใดถูกพบในบริเวณนั้นๆ

ค่า 1-6 ระบุผู้เล่น

ค่า 7 ไม่ถูกใช้งาน

2.2.5.3 ส่วนติดต่อกับผู้ใช้ในการติดตามโครงร่าง (NUI Skeletal Tracking)

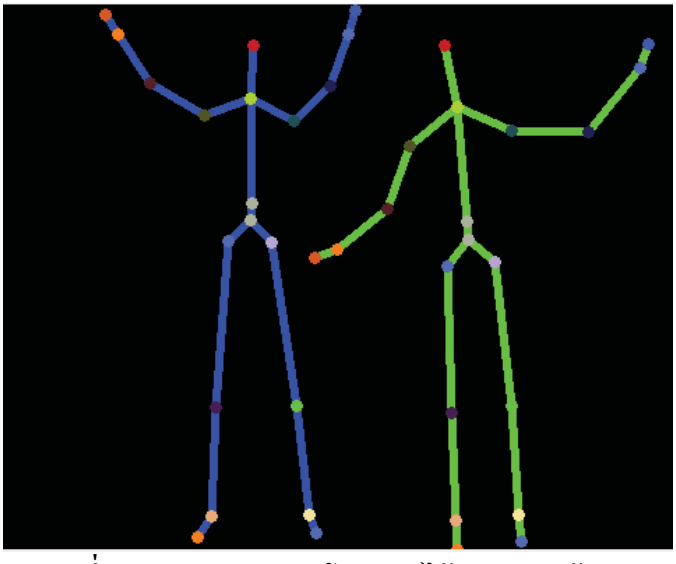
เอ็นยูไอโครงร่างเอพีไอ (NUI Skeleton API) ทำหน้าที่ให้ข้อมูลตำแหน่งปัจจุบันของผู้ใช้งานสามารถตรวจจับผู้ใช้งานได้จำนวน 2 คนประกอบไปด้วยชุดของข้อมูลตำแหน่งโครงร่างแต่ละจุดจำนวน 20 จุดที่ประกอบเป็นข้อต่อต่างๆของผู้ใช้ดังแสดงในรูป



รูปที่ 2.6 ตำแหน่งข้อต่อของโครงร่างที่สัมพันธ์กับร่างกายมนุษย์

2.2.5.4 การติดตามโครงร่าง (Skeletal Tracking)

การประมวลผลเพื่อหาโครงร่างสามารถติดตามตัวผู้เล่นภายในมุมมองของเซ็นเซอร์กล้องได้จำนวนผู้เล่นมากที่สุดจำนวน 2 ผู้เล่นเซ็นเซอร์กล้องยังสามารถตรวจจับผู้เล่นได้อีกจำนวน 4 ผู้เล่นภายในมุมมองของเซ็นเซอร์กล้องเมื่อผู้เล่นที่ถูกติดตามออกไปจากมุมมองของเซ็นเซอร์หรือได้รับข้อมูลของโครงร่างไม่ครบจะประมวลผลเพื่อติดตามโครงร่างจำนวน 2 โครงร่างแรกที่สามารถติดตามได้ขึ้นมาแทน



รูปที่ 2.7 สามารถติดตามโครงร่างได้จำนวน 2 ผู้เล่น

การส่งข้อมูลโครงร่างจะเป็นอาร์เรย์ของโครงสร้างข้อมูลของโครงร่างที่มีสถานะถูกติดตามอยู่เท่านั้นและจะประมวลผลเฟรมข้อมูลความลึกของโครงร่างนั้นๆ

ข้อมูลที่ได้จากการติดตามโครงร่าง

- สถานะปัจจุบันของแต่ละโครงร่าง (SkeletonTrackingState) ประกอบด้วย 3สถานะคือ NotTracked PositionOnly และ Tracked
- หมายเลขประจำตัวของโครงร่าง (TrackingID) ใช้ระบุตัวผู้เล่นเพื่อใช้จัดการกับอาร์เรย์ของโครงสร้างข้อมูลของแต่ละโครงร่างที่มีสถานะถูกติดตามอยู่ได้ถูกต้อง
- ตำแหน่งจุดศูนย์กลางมวล (Center of Mass) จะแสดงเฉพาะผู้เล่นที่สามารถตรวจจับได้ภายในมุมมองของเซ็นเซอร์กล้องแต่ไม่ถูติดตามเป็นโครงร่าง
- สำหรับผู้เล่นที่ถูกติดตามเป็นโครงร่างจำนวน2ผู้เล่นสามารถส่งกลับค่าข้อมูลของโครงร่างได้

- สำหรับผู้เล่นที่ไม่ถูกติดตามเป็นโครงร่างสามารถส่งกลับค่าข้อมูลได้เพียงแค่ค่าตำแหน่งของผู้เล่นและข้อมูลหมายเลขประจำตัวของโครงร่างเท่านั้น

2.2.5.5 Color Stream

ข้อมูลภาพสีมีสองระดับคุณภาพและสองรูปที่แบบที่แตกต่างกันระดับคุณภาพจะพิจารณาถึงวิธีการส่งข้อมูลจากเซ็นเซอร์ให้รวดเร็วที่สุดรูปที่แบบสีพิจารณาว่าสีของสตรีมข้อมูลภาพจะถูกเข้ารหัสเป็น RGB หรือ YUV

เซ็นเซอร์ใช้การเชื่อมต่อ USB เพื่อส่งข้อมูลและจัดเตรียมปริมาณของแบนด์วิดธ์ทางเลือกของคุณภาพข้อมูลภาพจะช่วยให้สามารถปรับแต่งวิธีการใช้แบนด์วิดธ์ภาพที่มีคุณภาพสูงส่งข้อมูลต่อเฟรมมากขึPนและมีอัตราการปรับปรุงน้อยกว่าภาพคุณภาพปกติทำให้ลดการสูญเสียคุณภาพของภาพจากกระบวนการบีบอัดเซ็นเซอร์ที่ส่งข้อมูลภาพสีที่ขนาด 1280x960 จะถูกบีบอัดและแปลงเป็น RGB หลังจากนั้นจะขยายข้อมูลก่อนส่งข้อมูลไปยังแอพพลิเคชันซึ่งการบีบอัดทำให้เกิดความเป็นไปได้ที่จะส่งข้อมูลเกี่ยวกับสีที่อัตราเฟรมสูงถึง 30 เฟรมต่อวินาทีแต่อาจเกิดการสูญเสียความเที่ยงตรงถูกต้องของข้อมูลเล็กน้อยจากอัลกอริทึมที่ใช้ข้อมูลสีที่ให้บริการมีสองรูปที่แบบดังตารางที่ 2.2 รูปที่แบบสีทั้งสองถูกคำนวณจากข้อมูลกล้องตัวเดียวกันดังนั้นข้อมูลทั้งสองประเภทจะแสดงในภาพเดียวกัน

ตารางที่ 2.2 ข้อมูลรูปที่แบบสี

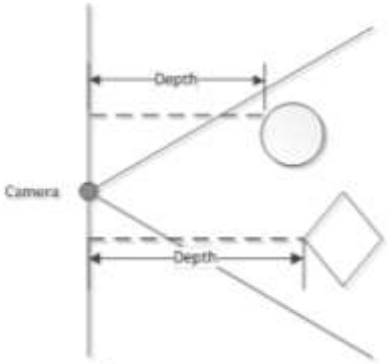
รูปแบบสี	คำอธิบาย
RGB	จำนวน 32 บิตรูปแบบสี X8R8G8B8 บิตแมปในพื้นที่RGB รูปแบบบิตสีโหมดนี้แสดงค่าเพียงแดงเขียวและน้ำเงินเท่านั้นหมายเลขที่ตามหลัง X หมายถึงไม่มีการนำเอาค่า 8 บิตซ้ายสุดของพิกเซลที่อยู่ในรูปที่แบบการแสดงผลนีPมาใช้งานสาเหตุที่ต้องนำเอารูปที่แบบการแสดงผลที่ประกอบไปด้วยค่าของ Xเนื่องจากมอนิเตอร์แสดงผลไม่สามารถแสดงส่วนที่โปร่งใสข้อมูล RGB มีความละเอียดที่ 640x480 15 เฟรมต่อวินาทีความละเอียด 640x480 30 เฟรมต่อวินาทีและความละเอียด 1280x960 ด้วยอัตรา 12 เฟรมต่อวินาที
YUV	จำนวน 16 บิตรูปแบบสี UYVY บิตแมปเมื่อแกมมาถูกเก็บในพื้นที่ YUVเทียบเท่ากับแกมมาRGB ในพื้นที่ RGB เนื่องจากสตรีม YUV ใช้ 16 บิตต่อพิกเซลจึงทำให้ใช้หน่วยความจำน้อยกว่าที่จะเก็บข้อมูลบิตแมปและมีการจัดสรรหน่วยความจำบัฟเฟอร์น้อยกว่าเพื่อเปิดสตรีมข้อมูล YUV มีความละเอียดที่640x480 ด้วยอัตรา 15เฟรมต่อวินาทีเท่านั้น

2.2.5.6 Coordinate Spaces

คินเน็คท์สามารถแสดงผลข้อมูลสีข้อมูลความลึกภาพเป็นค่าเฟรมต่อวินาทีข้อมูลแต่ละชนิดจะมีพื้นที่การแปลงพิกัดต่างกัดโดยมีส่วนต่อประสานโปรแกรมประยุกต์ (API)ทำหน้าที่ในการแปลงข้อมูลจากพื้นที่หนึ่งไปยังอีกพื้นที่หนึ่ง

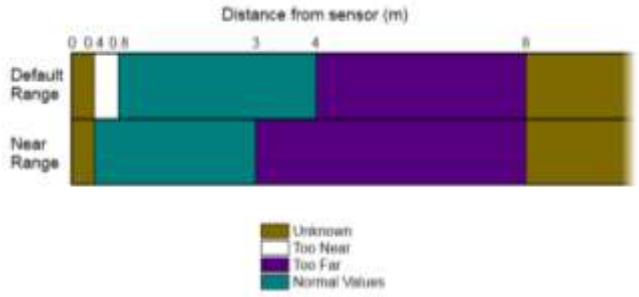
1) Color Space แต่ละเฟรมเซ็นเซอร์สีจะจับภาพสีของทุกอย่างที่เห็นได้ในฟิลล์ของมุมมองเซ็นเซอร์สีซึ่งเฟรมสร้างพิกเซลขึ้นจำนวนพิกเซลขึ้นอยู่กับขนาดของเฟรมแต่ละพิกเซลประกอบด้วยค่าสีแดงสีเขียวและสีน้ำเงินที่พิกัด (x, y) เฉพาะเจาะจง

2) Depth Space แต่ละเฟรมเซ็นเซอร์ความลึกจะจับภาพระดับสีเทาของทุกอย่างที่เห็นได้ในฟิลล์ของมุมมองเซ็นเซอร์ความลึกซึ่งเฟรมสร้างพิกเซลขึ้นจำนวนพิกเซลขึ้นอยู่กับขนาดของเฟรมแต่ละพิกเซลประกอบด้วยระยะทางคาร์ทีเซียนในหน่วยมิลลิเมตรจากระนาบกล้องไปที่วัตถุที่อยู่ใกล้ที่สุดที่พิกัด (x, y) เฉพาะเจาะจงดังรูปที่ 2.7



รูปที่ 2.7 ค่าสตรีมความลึก

เซ็นเซอร์ความลึกมีสองช่วงความลึกคือช่วงปกติและช่วงใกล้ดังรูปที่ 2.8 แสดงให้เห็นช่วงความลึกเซ็นเซอร์ในหน่วยเมตรช่วงปกติมีทั้งในคินเน็คท์สำหรับวินโดวส์และเอ็กซ์บ็อกซ์ 360 ส่วนช่วงใกล้จะใช้ได้เฉพาะในคินเน็คท์สำหรับวินโดวส์เท่านั้น



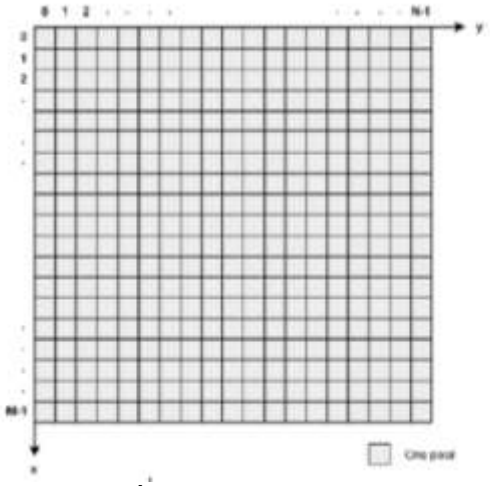
รูปที่ 2.8 ช่วงความลึกที่คินเน็คท์รับข้อมูลได้

2.3 การประมวลผลภาพ

การประมวลผลภาพ (Image Processing) คือการนำภาพมาประมวลผลโดยการแยกแยะส่วนที่ต้องการหรือสนใจออกจากพื้นหลังหรือสิ่งรบกวนต่างๆการเปลี่ยนแปลงค่าสีของภาพการหาขอบภาพการกรอง (Filter) การประมวลผลเชิงตัวเลขซึ่งขั้นตอนดังกล่าวนี้จะถูกนำมาคิดคำนวณและประมวลผลทางคอมพิวเตอร์การคิดคำนวณนั้นมีหลายวิธีซึ่งแต่ละวิธีก็มีประโยชน์แตกต่างกันไม่ว่าจะเป็นการคิดคำนวณในแต่ละพิกเซล (Pixel) หรือการคิดคำนวณเป็นบริเวณหลายๆจุดรวมกัน (Area) เช่นการดูลวดลาย (Pattern, Texture) การวิเคราะห์หารูปร่าง (Shape) และการวิเคราะห์แบบอื่นๆการประมวลผลภาพดิจิตอล (Digital Image Processing) เป็นสาขาที่กล่าวถึงเทคนิคและอัลกอริทึมที่ใช้การประมวลผลภาพที่อยู่ในรูปแบบดิจิตอล

2.4 ภาพดิจิตอล (Digital Image)

ภาพดิจิตอลเป็นฟังก์ชันของข้อมูล 2 มิติซึ่งมีการอ้างอิงค่าตามตำแหน่งของพิกเซลโดยค่าที่อ้างอิงในแต่ละตำแหน่งของพิกเซลส่วนใหญ่อยู่ในรูปของค่าความเข้มแสงและสีซึ่งค่าของพิกเซลสามารถแทนได้ในหลายรูปแบบตามประเภทของภาพ



รูปที่ 2.9 ตัวอย่างภาพดิจิตอล

2.4.1 องค์ประกอบของภาพดิจิตอล

ภาพดิจิตอลสามารถนิยามเป็นฟังก์ชันสองมิติf (x, y) โดยที่(x, y) เป็นพิกัดของภาพและแอมพลิจูดของf ที่พิกัด(x, y) ใดๆคือค่าความเข้มแสงของภาพ (Intensity) ที่ตำแหน่งนั้นๆเมื่อ(x, y) และแอมพลิจูดของf เป็นค่าจำกัด (Finite Value) กำหนดให้ภาพ$f(x, y)$ เป็นภาพดิจิตอลที่มีขนาดM แถวN คอลัมน์และมีพิกัดของจุดกำเนิด (Origin) ภาพที่ตำแหน่ง(x, y) = (0, 0) โดยจะสามารถเขียนสมการให้อยู่ในรูปแบบเมตริก (Matrix) ได้ดังนี้

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix} \quad (2.1)$$

ค่าในเมตริกแต่ละค่าจะเรียกว่าองค์ประกอบภาพหรือพิกเซลโดยตำแหน่ง(0, 0) จะอยู่ทางด้านซ้ายมือด้านบนของภาพการจัดลำดับตำแหน่งของจุดภาพจะเรียงจากซ้ายไปขวาในแต่ละเส้นภาพและการจัดลำดับของเส้นภาพจะเรียงจากบนลงล่างดังรูปที่ 2.10

รูปที่ 2.10 ตัวอย่างภาพดิจิตอลขนาด 290x195

องค์ประกอบของภาพดิจิตอลมีดังนี้

2.4.1.1 พิกเซล (Pixel)

พิกเซลหรือ(x, y)ซึ่งเป็นพิกัดของภาพเป็นส่วนที่เล็กที่สุดของภาพดิจิตอลถ้านำพิกเซลหลายๆพิกเซลมาเรียงต่อกันจะกลายเป็นภาพที่มีขนาดเท่ากับจำนวนพิกเซลด้านกว้างคูณ

จำนวนพิกเซลด้านสูงเช่นรูปขนาด 600x900 พิกเซลหมายความว่ารูปนี้มีขนาดกว้าง 600 พิกเซล และสูง 900 พิกเซล

2.4.1.2 ค่าความเข้มแสงของภาพ

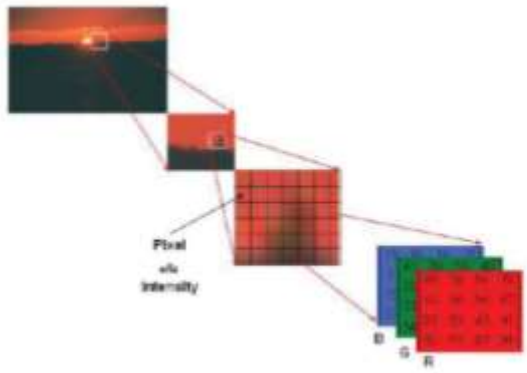
ค่าความเข้มแสงของภาพf(x, y)ในแต่ละจุดพิกเซลนั้นจะมีค่าตัวเลขกำกับไว้เพื่อบอกค่าระดับความเข้มแสงโดยใน 1 พิกเซลจะเป็นสีใดเพียงสีหนึ่งเท่านั้นจะมีสีอื่นไม่ได้ค่าความเข้มแสงของภาพดิจิตอลนั้นจะมีความแตกต่างกันไปตามประเภทของภาพดิจิตอล

2.4.2 ประเภทของภาพดิจิตอล

ภาพดิจิตอลมีอยู่หลากหลายประเภทเช่นภาพไบนารี่ภาพระดับสีเทาภาพดัชนีและภาพสี โดยภาพแต่ละประเภทนั้นจะมีค่าความเข้มแสงของภาพที่แตกต่างกัน

2.4.2.1ภาพสี (Color Image or RGB Image)

ภาพสีหรือภาพอาร์จีบี (RGB) คือภาพที่ค่าในแต่ละพิกเซลจะประกอบด้วยค่าของสีแดงสีเขียวและสีน้ำเงินโดยที่ค่าอาร์จีบีจะประมาณตามตัวรับสีในตาของมนุษย์และถูกใช้ในการแสดงภาพคอมพิวเตอร์และสแกนเนอร์หากภาพอาร์จีบีโดยปกติเป็น 24 บิตแต่ละสี (สีแดงสีเขียวและสีน้ำเงิน) จะมี 8 บิตสามารถเก็บความละเอียดของค่าความเข้มของแต่ละสีได้ 28 หรือ256 ระดับมีค่าอยู่ระหว่าง 0 ถึง 255 หากเป็นภาพอาร์จีบี 48 บิตแต่ละสี (สีแดงสีเขียวและสีน้ำเงิน) จะมี 16 บิตและสามารถเก็บความละเอียดของค่าความเข้มของแต่ละสีได้ 216 หรือ 65536 ระดับมีค่าอยู่ระหว่าง 0 ถึง 65535



รูปที่ 2.11ตัวอย่างภาพสีอาร์จีบีและค่าในแต่ละพิกเซลของภาพอาร์จีบี

2.4.2.2 ภาพสีเทา (Intensity Image or Gray Image)

ค่าในแต่ละพิกเซลของภาพสีเทาคือค่าความเข้มของแสงณแต่ละตำแหน่งของพิกเซลซึ่งจะอยู่ในรูปของระดับของภาพสีเทา (Gray Level)ขั้นตอนการแปลงภาพสีให้เป็นภาพระดับเทาทำได้โดยแยกระดับสีแต่ละพิกเซลออกจากกันในรูปแบบสีอาร์จีบีจากนั้นนำค่าสีอาร์จีบีมาเข้าสู่สมการเพื่อคำนวณหาค่าสีเทาและนำค่าที่ได้ไปแทนที่จุดพิกเซลเดิมภาพเกรย์สเกลหรือภาพระดับสีเทา โดยจะมีระดับความเข้มของสีเท่าคือ 0-255 (8 bit)ภาพเกรย์สเกลเกิดจากการแปลงภาพสีRGBมาเป็นภาพGrayscale โดยใช้สูตรทางคณิตศาสตร์ดังนี้

$$Gray=0.299\times R+0.587\times G+0.114\times B \qquad (2.2)$$

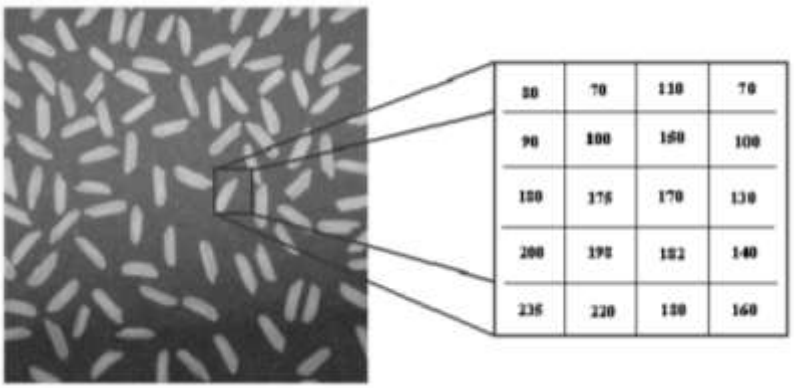
Gray = ค่าความเข้มของสีเทาโดยจะมีค่าระหว่าง 0-255

R = ค่าความเข้มของสีแดงโดยจะมีค่าระหว่าง 0-255

G = ค่าความเข้มของสีเขียวโดยจะมีค่าระหว่าง 0-255

B = ค่าความเข้มของสีน้ำเงินโดยจะมีค่าระหว่าง 0-255

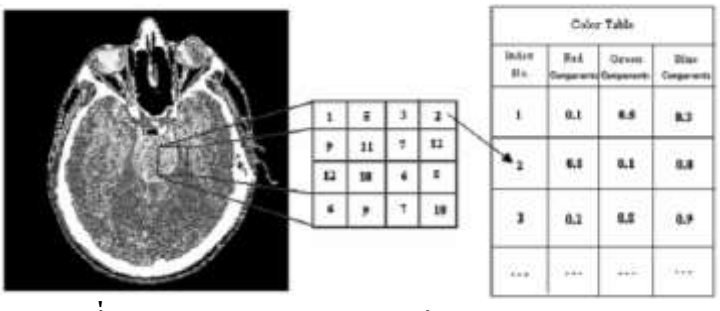
เช่นภาพระดับสีเทาขนาด 8 บิตสามารถแสดงค่าระดับสีเท่าได้ทัPงหมด 256ระดับและค่าที่เป็นไปได้คือ 0 ถึง 255ภาพระดับสีเทาหมายถึงภาพที่มีค่าความสว่างของแต่ละจุดภาพอยู่ในช่วงสีดำสีเทาเรื่อยไปจนถึงสีขาวซึ่งเป็นค่าความสว่างที่มากที่สุดสำหรับค่าของระดับสีเทานั้นปกติแล้วจะเป็นกำลังของ 2 ซึ่งโดยทั่วไปที่ใช้จะเป็น 8 บิต (256 ระดับ) หรืออาจใช้จำนวนบิตมากหรือน้อยกว่านี้ขึ้นอยู่กับความละเอียดของภาพที่ความต้องการว่ามากหรือน้อยเพียงใด



รูปที่ 2.12 ตัวอย่างภาพระดับสีเทาและค่าโทนสีเทาของแต่ละพิกเซล

2.4.2.3 ภาพดัชนี (Indexed Image)

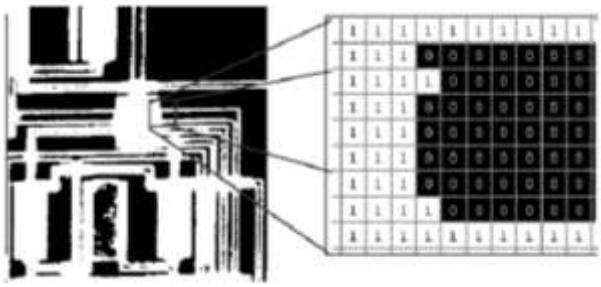
ภายในแต่ละจุดในภาพที่มีการทำดัชนีจะเก็บค่าดัชนีตัวเลข (Index Number) ซึ่งเป็นค่าที่ชี้ไปยังสีในตารางสีดังนั้นถ้าต้องการจะทราบสีในแต่ละจุดภาพจะต้องไปดูค่าสีที่ดัชนีอ้างอิงตรงกับค่าในจุดภาพ



รูปที่ 2.13 ตัวอย่างภาพระดับดัชนีและค่าของแต่ละพิกเซล

2.4.2.4 ภาพขาวดำหรือภาพไบนารี (Binary Image or Black and White Image)

ค่าในแต่ละพิกเซลของภาพขาวดำหรือภาพไบนารีจะใช้แค่ 1 บิตซึ่งจะมีค่าที่เป็นไปได้คือ 0 (สีดำ) และ 1 (สีขาว) เท่านั้น



รูปที่ 2.14 ตัวอย่างภาพไบนารี่และค่าในแต่ละพิกเซลของภาพไบนารี่